SCHOOL OF ENGINEERING - STI
SIGNAL PROCESSING LABORATORY LTS4
Vijayaraghavan Thirumalai and Pascal Frossard

CH-1015 LAUSANNE

Telephone: +4121 6932708
Telefax: +4121 6937600
e-mail: vijayaraghavan.thirumalai@epfl.ch

DISTRIBUTED REPRESENTATION OF GEOMETRICALLY CORRELATED IMAGES WITH COMPRESSED LINEAR MEASUREMENTS

Vijayaraghavan Thirumalai and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)

Signal Processing Laboratory LTS4 Technical Report

TR-LTS-2010-005

Dec 15th, 2010

Part of this work has been submitted to the IEEE Transactions on Image Processing.

This work has been supported by the Swiss National Science Foundation under grant 200021-118230.

Distributed Representation of Geometrically Correlated Images with Compressed Linear Measurements

Vijayaraghavan Thirumalai and Pascal Frossard
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4) , Lausanne, 1015 - Switzerland.
Email:{vijayaraghavan.thirumalai, pascal.frossard}@epfl.ch
Fax: +41 21 693 7600, Phone: +41 21 693 2708

Abstract

The distributed representation of correlated images is an important challenge in applications such as multi-view imaging in camera networks or low complexity video coding. This paper addresses the problem of distributed coding of images whose correlation is driven by the motion of objects or the positioning of the vision sensors. It concentrates on the problem where images are encoded with compressed linear measurements, which are used for estimation of the correlation between images at decoder. We propose a geometry-based correlation model in order to describe the correlation between pairs of images. We assume that the constitutive components of natural images can be captured by visual features that undergo local transformations (e.g., translation) in different images. These prominent visual features are estimated with a sparse approximation of a reference image by a dictionary of geometric basis functions. The corresponding features in the other images are then identified from the compressed measurements. The correlation model is given by the relative geometric transformations between corresponding features. We thus formulate a regularized optimization problem for the estimation of correspondences where the local transformations between images form a consistent motion or disparity map. Then, we propose an efficient joint reconstruction algorithm that decodes the compressed images such that they stay consistent with the quantized measurements and the correlation model. Experimental results show that the proposed algorithm effectively estimates the correlation between images in video sequences or multi-view data. In addition, the proposed reconstruction strategy provides effective decoding performance that compares advantageously to distributed coding schemes based on disparity or motion learning and to independent coding solution based on JPEG-2000.

Index Terms

Random projections, sparse approximations, motion estimation, disparity estimation, consistent reconstruction

I. Introduction

In recent years, vision sensor networks and video cellular phones have been gaining an ever increasing popularity that has been enforced by the availability of cheap semiconductor components. As these systems are operated with limited power, they require low complexity and power efficient algorithms for the processing and transmission of the visual information. Distributed processing becomes attractive in such settings since it involves a low complexity encoding stage that further permits to get rid of inter-sensor communication. In this case, the images captured by one or several image sensors are encoded independently but decoding jointly by a central decoder that exploits the underlying correlation. The computational complexity in the representation of the visual information is thus shifted from the encoder to the joint decoder.

In practice, the camera commonly acquires the image and then performs compression to reduce the transmission rate. Instead of acquiring the entire image, one could directly take the compressed data in the form of linear measurements, and the underlying signal can be reconstructed if it is sparse in a particular basis (e.g., DCT, Wavelet) [1], [2]. Such scheme allows for low complexity acquisition that consists in computing inner products with a random projection matrix, instead of acquiring the entire image. Hence, it is advantageous to merge the distributed processing and the image acquisition based on random projections, so that it results in a very simple encoding stage. One of the most important and challenging tasks in such a scenario is to estimate the correlation between the images (in terms of dense motion or disparity field) captured by different sensors or video cameras, so that the information can be efficiently processed or coded. When classical block-based motion estimation is performed, it is generally not possibly to efficiently capture the true geometry of the scene, which is key to an effective joint decoding of the correlated images.

In this paper, we consider the problem of finding an efficient distributed joint representation for a pair of correlated images, where the common objects in different images are displaced due to the view point change or motion of the scene objects. In particular, we are interested in computing a joint representation when the images are given under the form of few quantized linear measurements. We propose to model the correlation between images as the geometric transformation of visual features,

This work has been partly supported by the Swiss National Science Foundation, under grant 200021-118230.

rather than restricting ourselves to block-based translation correlation model. We first compute the most prominent visual features in a reference image and approximate them with geometric functions drawn from a parametric dictionary. Then we formulate an optimization problem whose objective is to compute the corresponding features in the compressed image along with the relative geometric transformations. We add a regularization constraint in order to ensure that the estimated motion (or disparity) field is consistent and corresponds to the actual motion of visual objects. The resulting correlation model is then used in a new joint reconstruction algorithm for computing an effective approximation of the correlated images. The joint decoding is cast as an optimization problem that includes a penalty term in order to enforce that the reconstructed image is consistent with the quantized measurements. We show by experiments that the proposed algorithm computes a good estimation of the motion or disparity field between the pair of images. In particular, the results confirm that dictionary based on geometric basis functions permits to capture the correlation more efficiently than a dictionary built on patches or blocks from the reference image [3]. In addition, we show that the estimated correlation model can be used to reconstruct the compressed image by motion (or disparity) compensation. Such reconstruction strategy permits to outperform DSC scheme based on disparity learning [4], [5] and independent coding scheme based on JPEG-2000 in terms of rate-distortion (RD) performance. Finally, the experiments outline the benefit of the consistent reconstruction penalty term in the joint reconstruction algorithm, where it proves to be very effective in increasing the decoding quality of the compressed images.

The rest of the paper is organized as follows. Section II briefly overviews the related work in signal processing based on random projections. The geometric based correlation model used in our framework is presented in Section III. Section IV describes the proposed regularized energy model and the correlation estimation algorithm, while the consistent reconstruction algorithm is presented in Section V. Experimental results in multi-view imaging and distributed video coding applications are given in Section VI.

II. Related work

We present in this section a brief overview of the related work in distributed image coding where we mostly focus on simple sensing solutions based on linear measurements. In recent years, signal acquisition based on random projections has actually received a significant attention in many applications like medical imaging, compressive imaging or sensor networks. Donoho [1] and Candes *et al.* [2] have shown that a small number of linear measurements may contain enough information to reconstruct a signal, as long as it has sparse representation in a basis that is incoherent with the sensing matrix [6]. These ideas have been applied to image acquisition [7], [8], [9] and later extended to video sequences [10], [11], [12].

The key in effective distributed representation certainly lies in the definition of good correlation models. Duarte *et al.* [13], [14] have proposed different correlation models for the distributed compression of correlated signals from linear measurements. In particular, they introduce three joint sparsity models in order to exploit the inter-signal correlation in the joint reconstruction. These three sparse models are respectively described by (i) JSM-1, where the signals share a common sparse support plus a sparse innovation part specific to each signal, (ii) JSM-2, where the signals share a common sparse support with different coefficients, and (iii) JSM-3 with a non-sparse common signal with individual sparse innovation in each signal. These correlation models permit a joint reconstruction with a reduced sampling rate or equivalently a smaller number of measurements compared to independent reconstruction for the same decoding quality. The sparsity models developed in [13] have then been applied for distributed video coding [15], [16] with random projections. The scheme in [15] used a modified Gradient projection sparse algorithm [17] for the joint signal reconstruction. The authors in [16] have proposed a distributed compressive video coding scheme based on the sparse recovery with decoder side information. In particular, the prediction error between the original and side information frames is assumed to be sparse in a particular orthonormal basis (e.g., Wavelet basis). An other distributed video coding scheme has been proposed in [3], which relies on an inter-frame sparsity model. A block of pixels in a frame is assumed to be sparsely represented by linear combination of the neighboring blocks from the decoded key frames. In particular, an adaptive block-based dictionary is constructed from the previously decoded key frames and eventually used for signal reconstruction. Finally, iterative projection methods are used in [18], [19] in order to ensure a joint reconstruction of correlated images that are sparse in a dual tree wavelet transform basis and at the same time consistent with the linear measurements in multi-view settings.

In multi-view imaging or distributed video coding, the correlation is explained by the motion of objects or view point change. Block-based translation models that are commonly used for correlation estimation fail to efficiently capture the geometry of scene objects. This results in poor correlation model, especially with low resolution images. Furthermore, most of the above mentioned schemes (except [3]) assume that the signal is sparse in a particular orthonormal basis (e.g., DCT or Wavelet). This is also the case of the JSM models above, which cannot be used to relate the scene objects by means of a local transform and unfortunately fail to provide an efficient joint representation of correlated images at lower rate. It is more generic to assume the signals to be sparse in a redundant dictionary, which allows greater flexibility in the design of the representation vectors. The most prominent geometric components in the images can be captured efficiently by dictionary functions. Then, the correlation can be estimated by comparing the most prominent features in different images. Few works have been reported in the literature for the estimation of a correlation model using redundant structured dictionaries in multi-view [20] or video applications [21]. However, these works do not construct the correlation model from the linear measurements. Rauhut *et al.* [22] extend the

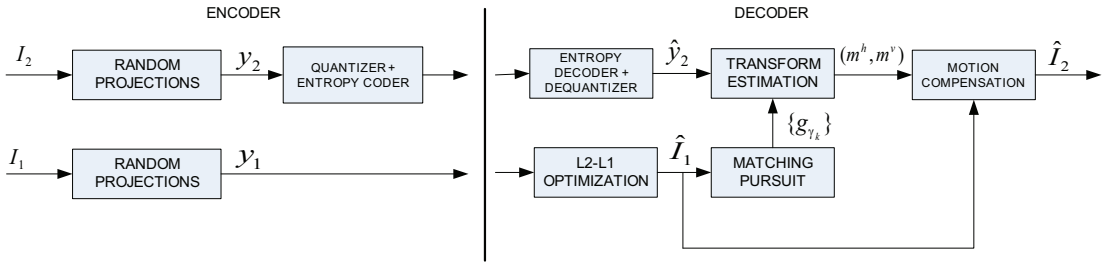


Fig. 1. Schematic representation of the proposed scheme. The images I_1 and I_2 are correlated through displacement of scene objects, due to view point change or motion of scene objects.

concept of signal reconstruction from linear measurements using redundant dictionaries. However, this has not been extended to distributed scenarios. We rather focus here on estimating the correlation from the random projections. The correlation model is built using the geometric transformations captured by the structured dictionary which leads to a good estimation of motion or disparity.

Finally, the distributed schemes based on compressed measurements (except [3]) usually fail to estimate the actual number of bits for the video representation and hence cannot be applied directly in practical coding applications. Quantization and entropy coding of the measurements in compressed sensing is actually an open research problem due to the two following reasons: (i) the reconstructed signal from quantized measurements does not necessarily satisfy the consistent reconstruction property [23]; (ii) the entropy of the measurements is usually large so that the coding performance in imaging applications is unsatisfactory [24]. Hence it is essential to develop adapted quantization techniques and reconstruction algorithms that reduce the distortion in the reconstructed signal, such as [25], [26]. The authors in [27], [28] have also studied the asymptotic reconstruction performance of the signal under uniform and non-uniform quantization schemes, and they have shown that the non-uniform quantization schemes usually give smaller distortion in the reconstruction signal, comparing to uniform quantization schemes. Recently, optimal quantization strategy for the random measurements has been designed based on distributed functional scalar quantizers [29]. In this paper, we use a simple quantization strategy with consistent reconstruction constraints in the joint decoding of correlated images, in order to illustrate the potential of low complexity sensing solutions in multi-view or distributed video coding applications.

III. Distributed coding with linear measurements

A. Framework

We consider a framework where a pair of images I_1 and I_2 represent a scene at different time instants or from different viewpoints; these images are correlated through the motion of visual objects. These images are represented by linear measurements that correspond to the projection of the image pixel values on a random set of coding vectors. They are then transmitted to a joint decoder that estimates the relative motion or disparity between the received signals and jointly reconstructs the images. The framework is illustrated in Fig. 1.

We focus on the particular problem where one of the images serves as a reference for the correlation estimation and the decoding of the second image. While this image could be encoded with any coding algorithm (e.g., JPEG-2000), we choose in this work to represent the reference image I_1 by random linear measurements $y_1 = \psi\ I_1$ with a projection matrix ψ. The measurements are used by the decoder to reconstruct an approximation $\hat{I}_1$ using a convex optimization algorithm under the assumption that I_1 is sparse in particular basis (e.g., a Wavelet basis) [9]. Next, we concentrate on the independent coding and joint decoding of the second image, where the first image serves as side information. The second image I_2 is also projected on a random matrix ψ to generate the measurements $y_2 = \psi\ I_2$. The measurements y_2 are quantized with a uniform quantization algorithm for the sake of simplicity at encoder; non-uniform quantization schemes are usually complex and demand the transmission of the codebook to the decoder. Finally, the bit rate is estimated by encoding the quantized linear measurements with an entropy coder (e.g., Arithmetic coder).

The joint decoder first computes the sparse approximation of the image $\hat{I}_1$ using the functions in a parametric dictionary of geometric functions. Such an approximation captures the most prominent geometrical features that represent the visual information in the image $\hat{I}_1$. The joint decoder then performs de-quantization and entropy decoding of the second image to form the measurement vector $\hat{y}_2$ (see Fig. 1). This measurement vector is used to estimate the relative transformation between the images I_1 and I_2 and also for reconstructing the second image when the first image serves as side information. Given the most prominent geometrical features in the image $\hat{I}_1$, we estimate the corresponding features in the second image I_2 whose visual information is given in terms of quantized linear measurements $\hat{y}_2$. In particular, the corresponding visual features in both images are then related using a geometry based correlation model, where the correspondences are defined under translational motion constraints. We generate the horizontal $\mathbf{m}^h$ and vertical $\mathbf{m}^v$ components of the dense motion field from the translation motion of the visual features between both the images. This motion information $(\mathbf{m}^h, \mathbf{m}^v)$ is further used

to reconstruct the compressed image $\hat{I}_2$ from the reference image $\hat{I}_1$. We further ensure a consistent reconstruction of $\hat{I}_2$ by explicitly considering the quantized measurements $\hat{y}_2$ during the reconstruction. Before getting into the details of the joint reconstruction algorithm, we describe the sparse image approximation algorithm and the geometry-based correlation model built on a parametric dictionary.

B. Sparse Image Approximation

We discuss here the sparse approximation using geometric basis functions in a structured dictionary, which helps to build the geometric correlation model between the images. We propose to represent the images by a sparse linear expansion of geometric function g_γ taken from a parametric and overcomplete dictionary $\mathcal{D} = \{g_\gamma\}$. The geometric function g_γ in the dictionary $\mathcal{D}$ is usually called *atom*. The dictionary is constructed by applying a set of geometric transformations to a generating function g. These geometric transformations can be represented by a family of unitary operators $U(\gamma)$, so that the dictionary takes the form $\mathcal{D} = \{g_\gamma = U(\gamma)g, \gamma \in \Gamma\}$ for a given set of transformation indexes Γ. Typically this transformation set consists of scaling s_x, s_y, rotation θ, and translation t_x, t_y operators, defined as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1/s_x & 0 \\ 0 & 1/s_y \end{bmatrix} \begin{bmatrix} cos\ \theta & sin\ \theta \\ -sin\ \theta & cos\ \theta \end{bmatrix} \begin{bmatrix} x - t_x \\ y - t_y \end{bmatrix}$$

where x, y defines the image coordinate. Thus, each of the transformation or equivalently each atom in $\mathcal{D}$ is indexed by five parameters.

We can then write the linear approximation of the reference image $\hat{I}_1$ with functions in $\mathcal{D}$ as

$$\hat{I}_1 \approx \sum_{k=1}^{N} c_k\ g_{\gamma_k}, \tag{1}$$

where $\{c_k\}$ are the set of N coefficients, and the number of atoms N used in the approximation of $\hat{I}_1$ is usually much lesser the number of atoms in $\mathcal{D}$. It should be noted that in our framework we are interested in approximating the reconstructed version of the first image $\hat{I}_1$ using the functions in the geometric dictionary D.

Given the dictionary $\mathcal{D}$, finding the sparse approximation of the image using N atoms as given in Eq. (1), is usually a NP hard problem. Even though the problem is NP hard, several suboptimal algorithms exists in the literature (e.g., basis pursuit, matching pursuit, etc.) in order to find the capture the most salient and prominent visual features in the images, in reasonable computational time. In this work we have chosen matching pursuit [30], which greedily pick up the N atoms $\{g_{\gamma_k}\}$ that best match the image $\hat{I}_1$ [31].

C. Joint Correlation Model

We propose to model the correlation between the images by the relative transformation of prominent visual features in both images that are captured by geometric functions from a structured dictionary. The correlation model is expected to offer better performance for a joint decoding strategy compared to independent approximation of each image. We describe briefly the correlation model between the images I_1 and I_2 in the rest of this section. For more details, we encourage the reader to refer [20], [21].

In this work, we arbitrarily choose I_1 as the reference image. We first compute the set of functions $\{g_{\gamma_k}\}$ that form the sparse approximation of the reconstructed version of the reference image $\hat{I}_1$ (see Eq. (1)). Under the assumption that the images are correlated via local geometric transforms of their principal components, the second image I_2 could be approximated with transformed version of the atoms used in the approximation of $\hat{I}_1$. We can thus write

$$I_2 \approx \sum_{k=1}^{N} c_k\ F^k(g_{\gamma_k}), \tag{2}$$

where $F^k(g_{\gamma_k})$ represents a local geometrical transformation of the atom g_{γ_k}. For consistency, we have described the correlation model given in Eq. (2) using the functions $\{g_{\gamma_k}\}$ that are picked from the reconstructed image $\hat{I}_1$. However, this correlation model holds well even if the atoms $\{g_{\gamma_k}\}$ are picked from the original image I_1. Due to the parametric form of the dictionary, the effect of F^k corresponds to a geometrical transformation of the atom g_{γ_k} that results in another atom in the same dictionary D. Therefore, it is interesting to note that the transformation F^k on g_{γ_k} boils down to a transformation $\delta\gamma$ of the atom parameters, i.e.,

$$F^k(g_{\gamma_k}) = U(\delta\gamma)g_{\gamma_k} = U(\gamma_k \circ \delta\gamma)g = g_{\gamma_k \circ \delta\gamma} = g_{\gamma'_k}. \tag{3}$$

The main challenge in the joint decoder consists in estimating the local geometrical transformation F^k for each of the atom g_{γ_k} in $\hat{I}_1$ from the compressed linear measurements $\hat{y}_2$. We formulate in the next section a regularized optimization problem in order to estimate F^k, or equivalently the relative motion or disparity between images I_1 and I_2.

IV. Correlation estimation from compressed linear measurements

Given the set of N atoms $\{g_{\gamma_k}\}$ that approximate the first image $\hat{I}_1$, the disparity or motion estimation problem consists in finding the corresponding visual patterns in the second image I_2, while the latter is given only by compressed random measurements $\hat{y}_2$. This is equivalent to find the correlation between both images with the joint sparsity model based on local geometrical transformations as described in Section III-C. We describe here the proposed regularized optimization framework that estimates the correlation between the images I_1 and I_2.

A. Regularized energy function

We are looking for a set of N atoms in I_2 that corresponds to the N visual features $\{\gamma_k\}$ selected in the first image. We denote this set by Λ, where $\Lambda = (\gamma'_1, \gamma'_2, ...\gamma'_N)$ for some $\gamma'_k \ \forall \ k$, $1 \leq k \leq N$. We propose to select this set of atoms in a regularized energy minimization framework as a trade-off between the set that well approximate I_2 and the set that results in smooth local transformations between both images. The energy model E proposed in our scheme is expressed as

$$E(\Lambda) = E_d(\Lambda) + \alpha_1 E_s(\Lambda), \tag{4}$$

where E_d and E_s represent the data term and smoothness term respectively, and α_1 is the regularization constant that balances the data and smoothness term. The solution to the correlation estimation is given by the set of N atom parameters Λ^* that minimizes the energy E, i.e.,

$$\Lambda^* = arg \min_{\Lambda \in S} E(\Lambda) \tag{5}$$

where S represents the search space given by

$$S = \{(\gamma'_1, \gamma'_2, ...\gamma'_N) \mid \gamma'_k = \gamma_k + \delta\gamma, 1 \leq k \leq N, \delta\gamma \in \mathcal{U}\}. \tag{6}$$

where $\mathcal{U} \subset \mathbb{R}^5$, and $\mathcal{U}$= $[-\delta t_x;\ \delta t_x] \times [-\delta t_y;\ \delta t_y] \times [-\delta\theta;\ \delta\theta] \times [-\delta s_x;\ \delta s_x] \times [-\delta s_y;\ \delta s_y]$ where $\delta t_x, \delta t_y$, $\delta\theta$, $\delta s_x, \delta s_y$ determine the search window size corresponding to translation parameters t_x, t_y, rotation θ and scales s_x, s_y respectively.

We consider now that the transformation between images are mainly due to translation motion with pixel accuracy, which represents a good motion model for distributed video coding or multi-view images captured by neighbor cameras. We look for transformations F^k that involves translation in horizontal and vertical directions with pixel accuracy, but changes in the scaling or orientation are not considered, i.e., the set of transformations F^k in Eq. (2) is limited to translations. Thus the search space for the set of atom parameters for the approximation of I_2 is limited to

$$S = \left\{(\gamma'_1, \gamma'_2, ...\gamma'_N) \mid \gamma'_k = (t_x^k + \Delta_x, t_y^k + \Delta_y, \theta^k, s_x^k, s_y^k)\right\} \tag{7}$$

with $1 \leq k \leq N, \Delta_x, \Delta_y \in \mathbb{Z}$, $-\delta t_x \leq \Delta_x \leq \delta t_x$, $-\delta t_y \leq \Delta_y \leq \delta t_y\}$ and $(t_x^k, t_y^k, \theta^k, s_x^k, s_y^k)$ represents the parameters of the atoms in $\{g_{\gamma_k}\}$. The motion field is estimated from the translations between the different pairs of corresponding atoms in both images. Given a pair of corresponding atoms g_{γ_k} and $g_{\gamma'_k}$ in the images I_1 and I_2 respectively, we first calculate the mapping of each pixel $\mathbf{z} = (x, y)$ in g_{γ_k} to its corresponding pixel $\tilde{\mathbf{z}} = (\tilde{x}, \tilde{y})$ in $g_{\gamma'_k}$ using Eq. (1). This grid transformation $\mathbf{z}^{(k)} - \tilde{\mathbf{z}}^{(k)} = (x^{(k)} - \tilde{x}^{(k)}, y^{(k)} - \tilde{y}^{(k)})$ corresponds to the amount of local motion captured by the k^{th} pair of atoms g_{γ_k} and $g_{\gamma'_k}$. Using the similar process, the mapping is established for all the N atom pairs from the respective transform parameters γ_k and γ'_k. Then the grid transformations captured by all the N pairs of atom are fused together to estimate the dense motion field. For the given location $\mathbf{z}$, we first assign weights $\{w_{\mathbf{z}}^{(k)}\}$ based on the response of the k^{th} atom at the pixel location $\mathbf{z}$. Then the fusion process is simply implemented by choosing the most confident transformation or motion $\mathbf{z}^{(k)} - \tilde{\mathbf{z}}^{(k)}$ for the given location $\mathbf{z}$, from the set of transformations $\{\mathbf{z}^{(k)} - \tilde{\mathbf{z}}^{(k)}\}$ induced by the N atoms. Thus the horizontal and vertical components of the motion field at location $\mathbf{z}$, denoted as $\mathbf{m}^h(\mathbf{z})$ and $\mathbf{m}^v(\mathbf{z})$, are given by

$$(\mathbf{m}^h(\mathbf{z}), \mathbf{m}^v(\mathbf{z})) = (x^{(k')} - \tilde{x}^{(k')}, y^{(k')} - \tilde{y}^{(k')}) \tag{8}$$

where $k' = arg \max_{k=1,2,...N} w_{\mathbf{z}}^{(k)}$, and $w_{\mathbf{z}}^{(k)}$ is the response of the k^{th} atom at the location $\mathbf{z}$ i.e., $w_{\mathbf{z}}^{(k)} = g_{\gamma_k}(\mathbf{z}) = g_{\gamma_k}(x, y)$. The resulting motion corresponds to motion of objects in video sequence, or to disparity values in multi-view imaging. We describe below the two cost functions used in Eq. (4).

B. Data cost function

Given the set of N atom parameters $\Lambda = \{\gamma'_k\}$, the data cost function E_d measures the error between the measurements $\hat{y}_2$ and the orthogonal projection of $\hat{y}_2$ onto the columns spanned by Ψ_Λ, where $\Psi_\Lambda = \psi[g_{\gamma'_1}|g_{\gamma'_2}|.....|g_{\gamma'_N}]$. It turns out that the orthogonal projection operator $\mathcal{P}$ is given by $\mathcal{P} = \Psi_\Lambda \Psi_\Lambda^\dagger$, where $\Psi^\dagger$ represents the pseudo-inverse. Therefore the data term estimates the set of N atom parameters Λ that agrees best with the measurements $\hat{y}_2$. More formally, the data cost is computed using the following relation,

$$E_d(\Lambda) = \parallel \hat{y}_2 - \Psi_\Lambda \Psi_\Lambda^\dagger \hat{y}_2 \parallel_2. \tag{9}$$

The data cost function given in the Eq. (9) first calculates the coefficients $c = \Psi_\Lambda^\dagger \hat{y}_2$, and then measures the l_2 distance between the observation $\hat{y}_2$ and $\Psi_\Lambda c$. However, when the measurements are quantized the coefficient vector c fails to properly account for the error introduced by the quantization. The quantized measurements give only an approximate bin information and the actual measurement value could be any point in the quantization interval. Let $y_{2,i}$ be the i^{th} coordinate of the original measurement, and $\hat{y}_{2,i}$ be the corresponding quantized value. Since the joint decoder has only access to the quantized value $\hat{y}_{2,i}$ and not the original value $y_{2,i}$, the joint decoder only knows that the quantized measurements lies within the quantized interval, i.e., $\hat{y}_{2,i} \in \mathcal{R}_{\hat{y}_i} = (r_i \ r_{i+1}]$, where r_i and r_{i+1} defines the lower and upper bound of quantizer bin $\mathcal{Q}_i$. We propose to refine the data cost term by computing a coefficient vector $\tilde{c}$ as the best solution when considering all the valid measurement values in the quantization interval, i.e., $\tilde{y}_2 \in \mathcal{R}_{\hat{y}}$, where $\mathcal{R}_{\hat{y}}$ is the cartesian product of quantized region $\mathcal{R}_{\hat{y}_i}$ for each i^{th} coordinate in $\hat{y}_{2,i}$. The coefficients $\tilde{c}$ and $\tilde{y}_2$ can be jointly estimated by solving the following optimization problem,

$$(\tilde{c},\ \tilde{y}_2) = arg \min_{\tilde{c},\tilde{y}_2} \parallel \tilde{y}_2 - \Psi_\Lambda \tilde{c} \parallel_2,\ s.t.\ \tilde{y}_2 \in \mathcal{R}_{\hat{y}}. \tag{10}$$

It can be shown that the Hessian of the objective function $f = \parallel \tilde{y}_2 - \Psi_\Lambda \tilde{c} \parallel_2$ is positive semidefinite, i.e., $\nabla^2 f \succeq 0$, and hence the objective function f is convex. Also the region $\mathcal{R}_{\hat{y}}$ forms a closed convex set as each region $\mathcal{R}_{\hat{y}_i} = (r_i \ r_{i+1}]$, $\forall i$ forms a convex set. Henceforth the optimization problem given in the Eq. (10) is convex. The data cost term given in Eq. (9) can be modified with the estimated coefficients $\tilde{c}$ as

$$\tilde{E}_d(\Lambda) = \parallel \hat{y}_2 - \Psi_\Lambda \tilde{c} \parallel_2 . \tag{11}$$

C. Smoothness cost function

The goal of the smoothness term E_s is to penalize the atom transformations such that they result in coherent deformations of neighbors atoms. In other words, the atoms in a neighborhood are likely to undergo similar transformation F^k when the correlation between images is due to object or camera motion. Instead of penalizing directly the transformation F^k to be coherent for neighbor atoms, we propose to generate a dense motion (or disparity) field from the atom transformation and to penalize the motion (or disparity) field such that it is coherent for adjacent pixels. This regularization is easier to handle than a regular set of transformations F^k and directly corresponds to the physical constraints that underly the formation of correlated images.

Once the motion field has been estimated using Eq. (8), the smoothness cost E_s is computed using the following relation

$$E_s = \sum_{\mathbf{z},\mathbf{z}' \in \mathcal{N}} V_{\mathbf{z},\mathbf{z}'}, \tag{12}$$

where $\mathbf{z}, \mathbf{z}'$ are the adjacent pixel locations and $\mathcal{N}$ is the usual 2 pixel neighborhood. The term $V_{\mathbf{z},\mathbf{z}'}$ in Eq. (12) is defined as

$$V_{\mathbf{z},\mathbf{z}'} = \min\left(|\mathbf{m}^h(\mathbf{z}) - \mathbf{m}^h(\mathbf{z}')| + |\mathbf{m}^v(\mathbf{z}) - \mathbf{m}^v(\mathbf{z}')|, K\right) \tag{13}$$

The parameter K sets a maximum limit to the penalty, and thus helps to preserve the discontinuities in the motion field [32]. It should be noted that for large K the smoothness term E_s becomes simply the cumulative sum of the dissimilarities in the motion field at adjacent pixels $\mathbf{z}, \mathbf{z}'$. In this case, it is equivalent the total variation (TV) norm of the motion field, i.e., the l_1 norm of the motion field gradient.

D. Optimization algorithm

As described in Section IV-A, we are interested in capturing the translation motion (or disparity) of the objects in the scene with pixel accuracy. We approximate the transformation F^k to act only on the integer locations of the translational component (t_x, t_y) of the atom g_{γ_k}, and the action of F^k on the rotation plus the scales of atom are not considered. Thus we are interested to estimate the set of parameters Λ^* that minimize the energy function E given in Eq. (4) when the search space is limited to translations of atoms in $\{g_{\gamma_k}\}$, as given in Eq. (7).

One could use an exhaustive search on the entire parameter space S (see Eq. (7)) to solve the optimization problem in Eq. (5). However, the cost for such a solution is high, as the size of the search space S grows linearly with the number of atoms N and exponentially with the window size, i.e., $|S| = N^{((2\delta t_x+1)\times(2\delta t_y+1))}$. Alternatively, Dynamic programming can be used to search for the global minima in this problem that is highly non-convex in S. Dynamic programming is an optimization methodology that usually decomposes the complex problem into several overlapping subproblems. Each subproblem is solved one by one starting from the smallest subproblem, and the solution to the current subproblem is estimated based on the solution estimated in the previous subproblem. The final solution is estimated by back-tracing the solution estimated in each subproblem. For energy function minimization described in Eq. (5), it is however hard to identify proper subproblems due to the overlap between the N atoms in the image approximation.

We propose here a parametric free optimization algorithm to estimate the transformation F^k iteratively, by changing each of the N atom parameters γ_k by one increment in the parameter space. We focus on the search space that is given by perturbing the translational components t_x and t_y of each atom position by one unit, i.e., $t_x \pm 1$, $t_y \pm 1$ for each atom γ_k. We first initialize

the algorithm with zero motion, i.e., the set of atoms $\{g_{\gamma_k}\}$ generated from $\hat{I}_1$ are used in the first iteration, $\gamma'_k = \gamma_k, \forall k$ where $1 \leq k \leq N$, and the search space is S^0 is formed using

$$\begin{aligned} S^0 =&\{(\gamma'_1, \gamma'_2, ..., \hat{\gamma}'_k, ..., \gamma'_N) | \hat{\gamma}'_k = (t_x^k + j_1, t_y^k + j_2, \theta^k, s_x^k, s_y^k), \\ & 1 \leq k \leq N, j_1, j_2 \in \mathbb{Z}, -1 \leq j_1, j_2 \leq 1\} \subset S. \end{aligned} \tag{14}$$

We then calculate the energy E in Eq. (4) for the set of N atoms in the search space S^0. It can be easily shown that the size of the search space S^0 is at most $8N + 1$, i.e., $|S^0| = 8N + 1$. Once the energy E is computed for atoms in S^0, we select the parameters $\Lambda^0 = (\gamma_1^0, \gamma_2^0, ..., \gamma_N^0)$ corresponding to the minimum energy. Then a new search space S^1 is formed similarly to the definition in Eq. (14) with the current parameter solution Λ^0 as reference. Such a procedure is repeated by successively constructing a new search space S^i on the solution Λ^{i-1} from the previous iteration of the algorithm. The algorithm stops when convergence is attained (or till it reaches maximum number of iterations). The proposed algorithm is guaranteed to converge. Let E_0 be the initial energy, i.e., the energy corresponding to set of parameters $\gamma'_k = \gamma_k, \forall k$ where $1 \leq k \leq N$. If E_i is the minimal energy computed at step i of the algorithm, we clearly have $E_i \leq E_{i-1}$, as the search space S^i includes the best set of parameters Λ^{i-1} from the previous iterations. The energy decreases at every iteration till it reaches a local or global minima E_{min}. The proposed optimization scheme thus converges and provides a (suboptimal) solution with tractable computational complexity to the estimation of correlation between images. The algorithm which bears some resemblance to a gradient descent solution is summarized in Algorithm 1. Finally, the data cost in the 8^{th} line of the Algorithm 1 can be replaced by the robust data cost term $\tilde{E}_d$ as given in Eq. (11). We show later that the performance of our scheme can be improved by using the robust data cost term $\tilde{E}_d$.

Algorithm 1 Correlation estimation

1: Input N, α_1, K, δt_x, δt_y
2: Generate $\{g_{\gamma_k}\}$ from $\hat{I}_1$ s.t. $\hat{I}_1 \approx \sum_{k=1}^{N} c_k \; g_{\gamma_k}$
3: Initialize $\Lambda^{-1} = \{\gamma_k\}$
4: *repeat*
5: Generate index search space S^i based on Λ^{i-1} (with Eq. (14))
6: **for all** Parameter vectors Λ in S^i **do**
7: Compute the motion field
8: Compute the data term $E_d(\Lambda)$ with Eq. (9)
9: Compute the smoothness term $E_s(\Lambda)$ with Eq. (12)
10: Compute the global energy $E(\Lambda) = E_d(\Lambda) + \alpha_1 E_s(\Lambda)$
11: **end for**
12: $\Lambda^i = arg\min_{\Lambda \in S^i} E(\Lambda)$
13: Until convergence is reached

V. Consistent reconstruction by warping

Once the motion or disparity between the correlated images has been estimated as described in the previous section, one can simply reconstruct the second image by warping the reference image the reference view $\hat{I}_1$ using the estimated motion or disparity field. The resulting approximation $\hat{I}_2$ is however not necessarily consistent with the quantized measurements $\hat{y}_2$. In other words, the measurements corresponding to the projection of the image $\hat{I}_2$ on the sensing matrix Ψ are not necessarily equal to $\hat{y}_2$. This error might even be quite significant.

We propose to add a consistency term E_t in the energy model described in Eq. (4) in order to force consistency in the image reconstruction through warping with operator $\mathcal{W}_\Lambda$ using the computed motion (or disparity) field (see Fig. 1). In particular, the additional cost function E_t is defined as the l_2 norm error between the quantized measurements generated from the reconstructed image $\hat{I}_2 = \mathcal{W}_\Lambda(\hat{I}_1)$ and the measurements $\hat{y}_2$. The cost function E_t is written as

$$E_t(\Lambda) = \| \; \hat{y}_2 - \mathcal{Q}[\psi \hat{I}_2] \; \|_2 = \| \; \hat{y}_2 - \mathcal{Q}[\psi \mathcal{W}_\Lambda(\hat{I}_1)] \; \|_2 \tag{15}$$

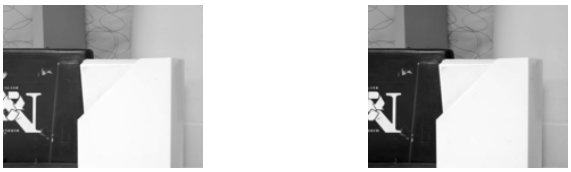
where $\mathcal{Q}$ is the quantizer. It should be noted that when the measurements are not quantized the consistency term reads

$$\tilde{E}_t(\Lambda) = \| \; y_2 - \psi \mathcal{W}_\Lambda(\hat{I}_1) \; \|_2 \tag{16}$$

We then merge the three cost functions E_d, E_s and E_t with regularization constants α_1 and α_2 in order to form a new energy model E_R for consistent reconstruction expressed as

$$E_R(\Lambda) = E_d(\Lambda) + \alpha_1 E_s(\Lambda) + \alpha_2 E_t(\Lambda). \tag{17}$$

In order to solve Eq. (17) or equivalently to estimate the correlation model that leads to consistent reconstruction, we propose to use the same optimization method as the one described in Section IV-D. We modify the objective functions to include the



(a) (b)

Fig. 2. Plastic image set (a) View point 1 (I_1) (b) View point 2 (I_2)

consistency term, and we iteratively look for the best atoms parameters sets by adapting successively the translation parameters. Again, the algorithm is guaranteed to converge to a local or global minimum with a moderate computational complexity. The consistent reconstruction algorithm is summarized in Algorithm 2. Finally, the data cost term in the 9^{th} line of Algorithm 2 can be replaced with robust data term $\tilde{E}_d$ to provide robustness to quantization errors.

Algorithm 2 Consistent reconstruction

1: Input N, α_1, K, δt_x, δt_y
2: Generate $\{g_{\gamma_k}\}$ from $\hat{I}_1$ s.t. $\hat{I}_1 \approx \sum_{k=1}^{N} c_k \; g_{\gamma_k}$
3: Initialize $\Lambda^{-1} = \{\gamma_k\}$
4: *repeat*
5: Generate index search space S^i based on Λ^{i-1} (with Eq. (14))
6: **for all** Parameter vectors Λ in S^i **do**
7: Compute the motion field
8: Warp the reference image $\hat{I}_1$ using motion field, $\mathcal{W}_\Lambda(\hat{I}_1)$
9: Compute the data term $E_d(\Lambda)$ with Eq. (9)
10: Compute the smoothness term $E_s(\Lambda)$ with Eq. (12)
11: Compute the consistency term $E_t(\Lambda)$ with Eq. (15)
12: Compute the global energy $E_R(\Lambda) = E_d(\Lambda) + \alpha_1 E_s(\Lambda) + \alpha_2 E_t(\Lambda)$
13: **end for**
14: $\Lambda^i = arg\min_{\Lambda \in S^i} E_R(\Lambda)$
15: Until convergence is reached

We discuss now briefly the computational complexity of the algorithm for image reconstruction, which can basically be divided into two stages. The first stage finds the most prominent features in the reference image using sparse approximations in a structured dictionary, and the second stage estimates the transformation for all the features in the reference image by solving a regularized optimization problem. Even if the decoder can afford computational complexity in our framework, both stages might be appear complex. However, the computational complexity of the decoder can be reduced significantly using a tree-structured dictionary for the approximation of the reference image [33]. Alternatively, a block-based dictionary can be used, and transformations are then computed for each block. Experiments show however that this comes at a price of a performance penalty in the reconstruction quality. It is clear that the decoding scheme proposed here offers high flexibility with a tradeoff between complexity and performance. For example, one might decide to use the simple data cost E_d even when the measurements are quantized. This leads to a simpler scheme but to reduced reconstruction quality. Overall, our framework offers a very simple encoding stage with image acquisition based on random linear projections and the computational burden is shifted to a joint decoder that can trade-off complexity and performance.

VI. Experimental results

A. Overview

We analyze the performance of the correlation estimation and image reconstruction algorithms in multi-view imaging and distributed video coding applications. The images are captured by random linear projections using the scrambled block Hadamard transform with block size 8 [9]. The reference image is then reconstructed as the solution of a $\ell_1 - \ell_2$ regularization problem using a GPSR algorithm [17]. In order to compute a sparse approximation of the reference image at decoder, we use a dictionary that is constructed using two generating functions, as explained in [31]. The first one consists of 2D Gaussian functions to capture low frequency components. The second function represents Gaussian in one direction, and the second derivative of 2D gaussian in the orthogonal direction to the capture edges. The discrete parameters of the functions in the dictionary are chosen as follows. The translation parameters t_x and t_y take any positive value and cover the full width and

(a) (b)

Fig. 3. Sawtooth image set (a) View point 1 (I_1) (b) View point 5 (I_2)

(a) (b) (c) (d) (e)

Fig. 4. Plastic image set (a) Ground truth disparity field $\mathbf{M}^h$ between views 1 and 2 (b) Disparity field with Alg. 1 (c) Error in the depth map with Alg. 1 (DE = 0.10). (d) Disparity field with Alg. 2 (d) Error in the depth map with Alg. 2 (DE = 0.05). [8870 quantized measurements]

height of the image. Ten rotation parameters are used between 0 and π, with increments $\pi/18$. Five scaling parameters are equidistributed in the logarithmic scale from 1 to $N_1/8$ vertically, and 1 to $N_2/9.77$ horizontally, where $N_1 \times N_2$ is the size of the image.

The measurements y_2 corresponding to the second image I_2 is computed with the same sensing matrix ψ as the reference image. The measurements y_2 are quantized a uniform quantizer, and the bit rate is computed by encoding the quantized measurements using an Arithmetic coder. We report in this section the performance of the correlation estimation and analyze the influence of the measurement consistency term in the energy minimization constraint. We also analyze the influence of the quantization of the measurements for the second image. Then we study the reconstruction results as a function of the measurement rate or the coding rate of the second image. We also compare the performance of our disparity estimation algorithms with stochastic optimization algorithm based on simultaneous perturbation (SPSA) [34]. Finally, we compare the rate-distortion performance in the coding of the second image to state-of-the-art solutions for independent or distributed image coding.

B. Multi-view image coding

We first study the performance of our distributed image representation algorithms in a multi-view imaging framework. We use two image datasets, namely *Plastic* (see Fig. 2) and *Sawtooth* (see Fig. 3) [1]. These datasets have been captured by a camera array where the different viewpoints are arranged uniformly arranged on a line. As this corresponds to translating the camera along one of the image coordinate axis, the disparity estimation problem becomes a one-dimensional search problem and the smoothness term in Eq. (8) is simplified accordingly. The images are downsampled to a resolution 144×176 using bilinear filters. We carry out experiments using the views 1 and 5 for Sawtooth image set, and views 1 and 2 for Plastic image set. The view point 1 is selected as the reference image I_1. Unless stated differently, it is encoded such that the quality of $\hat{I}_1$ is

[1]These image sets are available in http://vision.middlebury.edu/stereo/data/

(a) (b) (c) (d) (e)

Fig. 5. Sawtooth image set (a) Ground truth disparity field $\mathbf{M}^h$ between views 1 and 5 (b) Disparity field with Alg. 1 (c) Error in the depth map with Alg. 1 (DE = 0.092). (d) Disparity field with Alg. 2 (d) Error in the depth map with Alg. 2 (DE = 0.047). [8870 quantized measurements]

$(a) MSE: 102$ $(b) MSE: 558$ $(c) MSE: 44$ $(d) MSE: 633$

Fig. 6. Comparison of the warped image $\hat{I}_2$ w.r.t. I_2 and I_1 for the Plastic image set: (a) $1-|\hat{I}_2-I_2|$ with Alg. 1 (b) $1-|\hat{I}_2-I_1|$ with Alg. 1 (c) $1-|\hat{I}_2-I_2|$ with Alg. 2 (d) $1-|\hat{I}_2-I_1|$ with Alg. 2. [2534 quantized linear measurements].

approximately 33 dB w.r.t. I_1. Matching pursuit is then carried out on $\hat{I}_1$ by approximation with $N = 30$ and $N = 60$ atoms for Plastic and Sawtooth image set, respectively. The measurements are generally quantized using a 2-bit uniform quantizer. At decoder, the search for the geometric transformations F^k between images is carried out along the translational component t_x with window size $\delta t_x = 4$ pixels, and no search is consider along the vertical direction, i.e., $\delta t_y = 0$. Unless stated explicitly we use the data cost E_d given in Eq. (9) in Algorithms 1 and 2.

We first study the performance of the estimated disparity information and we show in Fig. 4 and Fig. 5 the estimated disparity field $\mathbf{m}^h$ from 8870 quantized measurements (i.e., a measurement rate of 35%) for both image sets respectively. The ground truth $\mathbf{M}^h$ is given in Fig. 4(a) and Fig. 5(a) respectively. The transformation F^k is estimated using the procedure described in Algorithm 1, and the resulting dense disparity fields are illustrated in Fig. 4(b) and Fig. 5(b). The Algorithm 1 gives a good estimation of the disparity map; in particular the disparity value is correctly estimated in the regions with texture or depth discontinuity. We could also observe that the estimation of the disparity field is however less precise in smooth regions as expected from feature-based methods. Fortunately enough, the wrong estimation of the disparity value corresponding to the smooth region in the images does not significantly affect the warped or predicted image quality [35]. Fig. 4(c) and Fig. 5(c) confirm such a distribution of the disparity estimation error and illustrate by white pixels the positions where the estimation error is larger than one. The disparity error DE is computed between the estimated disparity field $\mathbf{m}^h$ and ground truth $\mathbf{M}^h$ as $DE = \frac{1}{Z}\sum_{\mathbf{z}=(x,y)} \left\{|\mathbf{M}^h(\mathbf{z}) - \mathbf{m}^h(\mathbf{z})| \geq 1\right\}$ where Z represents the pixel resolution of the image [35]. We can see that the error in the disparity field is relatively high close the edges since crisp discontinuities cannot be accurately captured due to the scale and smoothness of the atoms in the chosen dictionary. The disparity information estimated by Algorithm 2 is presented in Fig. 4(d) and Fig. 5(d) and the corresponding errors in Fig. 4(e) and Fig. 5(e). We see that the addition of the consistency term E_t in the correlation estimation algorithm clearly improves the performance.

We propose a different illustration of the disparity estimation performance in Fig. 6. The dense disparity field $\mathbf{m}^h$ is used to warp the reference image $\hat{I}_1$ and the image thus reconstructed is represented by $\hat{I}_2$. We estimate the correlation between images in the Plastic dataset with Alg. 1 using 2534 quantized measurements (i.e., a measurement rate of 10%). We then warp the reference image to reconstruct an approximation $\hat{I}_2$ of the second image. We show in Fig. 6(a) and (b) the comparison between $\hat{I}_2$ and respectively I_2 and I_1, where white pixels represent a correct reconstruction. It is clear that $\hat{I}_2$ is closer to I_2 than I_1, which confirms that the proposed scheme capture the correlation between the images efficiently. The same comparisons are given in Fig. 6(c) and (d) when the Alg. 2 is used for correlation estimation. The results confirm that the addition of the consistency term again provides a more accurate disparity field since the warped image $\hat{I}_2$ gets quite close to the target image I_2.

We study now the rate-distortion performance of the proposed algorithms for the reconstruction of the image $\hat{I}_2$ in Fig. 7 for both datasets. We show the performance of the reconstruction by warping the reference image according to the correlation computed by Alg. 1 (without the measurement consistency term, i.e., $\alpha_2 = 0$ in Eq. (17)) and Alg. 2. We compare the rate-distortion performance to a distributed coding solution (DSC) based on LDPC encoding of the DCT coefficients, where the disparity field is estimated at the decoder using Expected Maximization (EM) principles [4]. The scheme is denoted as *Disparity learning* in the figures. Then, in order to demonstrate the benefit of geometric dictionaries, we also propose a scheme denoted as *block-based* that adaptively constructs the dictionary using blocks or patches in the reference image [3]. As described in [3], we construct a dictionary in the joint decoder from the reference image $\hat{I}_1$ using 8×8 blocks. The search window size is $\delta t_x = 4$ pixels along the horizontal direction. We then used the optimization scheme described in Alg. 2 to select the best block from the adaptive dictionary. In order to have a fair comparison, we encode the reference image I_1 similarly for both schemes (*Disparity learning* and *block-based*) with a quality of 33 dB (see Section III). Finally, we also provide the performance of a standard JPEG-2000 independent encoding of the image I_2. We first see by comparing the two proposed algorithms that the measurement consistency term greatly improves the decoding quality. Then, the results confirm that the proposed algorithms unsurprisingly outperform independent coding based on JPEG-2000, which outlines the benefits of the use of correlation in the decoding of compressed correlated images. At high rates, the performance of the proposed algorithms however tends to saturate as our model mostly handles the geometry and the correlation between images, but it is not able to

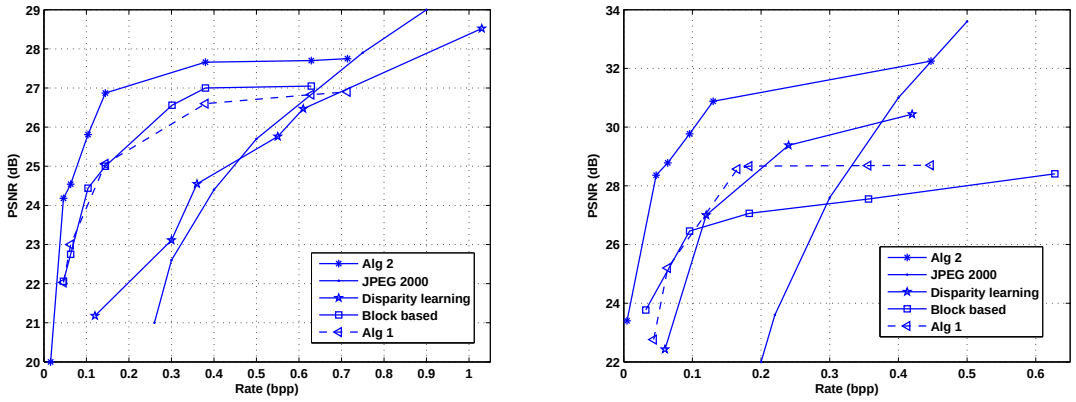


Fig. 7. Comparison of the RD performance of the proposed scheme w.r.t. DSC scheme [4], block based scheme [3] and independent coding solutions based on JPEG-2000. (a) Sawtooth image set (b) Plastic image set.

efficiently handle the fine details or texture in the scene due to the choice of dictionary functions. From Fig. 7(a) it is then clear that the reconstruction based on Alg. 1 outperforms the DSC coding scheme based on EM principles, for the Sawtooth image set. On the other hand for the Plastic image set (see Fig. 7(b)) our scheme performs very similar to the DSC scheme till rate 0.2 bpp, and after that our scheme performs around 2 dB lower than the DSC scheme. This is because, the Plastic scene mainly contains low frequency components or smooth regions, and in such regions it is hard to solve the correspondence problem or equivalently the estimation of the disparity [35]. However, when consistency with measurements is enforced in Alg. 2, we outperform DSC scheme based on disparity learning on both datasets. Finally, the experimental results also show that our schemes outperform the scheme based on block-based dictionary mainly because of the richer representation of the geometry and local transformations with the structured dictionary.

We further evaluate the performance of our constructive parameter search strategy in Alg. 2. We compare its performance to a global stochastic optimization method on SPSA [34] in terms of reconstruction performance and speed of convergence. We select a global optimization based on SPSA as it performs better than genetic algorithm [36], [37], and also gives better convergence rate than simulated annealing [37], [38], [39]. Similar to gradient descent algorithm, SPSA is an iterative algorithm that starts with an initial guess $\gamma'_k = \gamma_k, \ \forall k$ where $1 \leq k \leq N$, and computes an approximate noisy value of the gradient of the energy function E at the current solution $(\gamma'_1, \gamma'_2, ..., \gamma'_N)$, and updates the current solution $(\gamma'_1, \gamma'_2, ..., \gamma'_N)$. Maryak and Chin have studied the usage of SPSA as a global optimization and proved that SPSA under certain conditions can converge in probability to the global minima among multiple local minima [36]. We use a discrete SPSA algorithm where the gain sequence used in the gradient update is constant to the nearest integer [40], [41]. We examined the performance of the Sawtooth image set by setting $A = 0$, $\beta_2 = 0.95$, $a = 0.2$ and the sequence c_i is fixed to a constant $c = 1$, and the number of iterations, $iter = 2000$ in the SPSA algorithm, based on trial and error tests. We compare experimentally the speed of convergence between Alg. 2 and the SPSA algorithm. Fig. 8 shows the energy of the cost function as a function of the number of iteration in the disparity search algorithm for Sawtooth image set at measurement rate 3%. It is clear from the plot that Alg. 2 converges faster than the SPSA algorithm. Even if this is not a formal comparison in terms of computational complexity, it shows that parameter-free constructive algorithm fastly reaches a local minima, while global stochastic optimization may reach a better solution at the price of more computation and propser settings of multiple parameters. We also compare the RD performance of the reconstructed image $\hat{I}_2$ between the proposed and SPSA optimization algorithms. The reconstruction is done by warping the reference image based on the correlation estimation based on Alg. 2 and SPSA, respectively. Fig. 9 shows that SPSA performs slightly better than the proposed scheme due to a better estimation of the correlation. On the other hand the proposed scheme convergences faster than the SPSA algorithm, which leads to a trade-off between the speed of convergence and the RD performance. Similar observations have been made in our experiments for Plastic image set.

We now study the performance of Alg. 2 in different settings in terms of camera distances, quality of the reference image and quantization of the measurements. We first illustrate the rate-distortion performance of the proposed scheme for different images captured at various distances from the reference camera. In particular, we study the decoding quality of the images at viewpoints 3 and 5 in the Sawtooth dataset when the viewpoint 1 is used as the reference. Fig. 10 confirms that the performance is better when the correlation between images is stronger (the reference image is closer to viewpoint 3 than viewpoint 5). We further see that the coding performance is better than a state-of-the-art independent coding with JPEG-2000 or DSC based on disparity learning [4] when the correlation between images is high. We further study the influence of the quality of reference image $\hat{I}_1$ on the reconstruction performance. We use Alg. 2 to reconstruct $\hat{I}_2$ by warping when the reference image has been

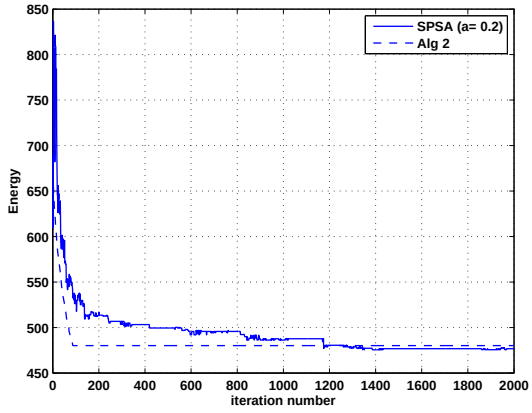


Fig. 8. Comparison of the speed of convergence between Alg. 2 and the SPSA optimization algorithm in the Sawtooth Image set. The experiments are carried out using SPSA parameters $A = 0$, $\beta_2 = 0.95$, and the sequence c_i is fixed to a constant $c = 1$. [Measurement rate = 3%].

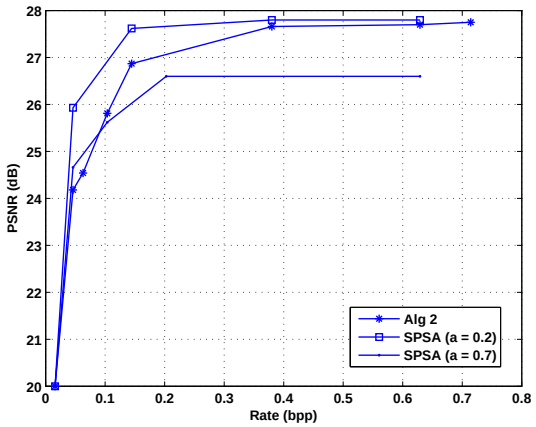


Fig. 9. Comparison of the rate-distortion performance in reconstructing $\hat{I}_2$ by warping for Alg. 2 and the SPSA optimization algorithm in the Sawtooth Image set. The experiments are carried out using SPSA parameters $A = 0$, $\beta_2 = 0.95$, and the sequence c_i is fixed to a constant $c = 1$.

encoded at different qualities (i.e., different measurement rates). Fig. 11 shows that the reconstruction quality $\hat{I}_2$ improves with the quality of the reference image $\hat{I}_1$, as expected. While the error in the disparity estimation is not dramatically reduced by improved reference quality the warping stage permits to provide more details in the representation of $\hat{I}_2$ when the reference is of better quality. Finally, we show the influence of the quantization rate on the rate-distortion performance in the reconstruction of $\hat{I}_2$ with Alg. 2. We have quantized the measurements $\hat{y}_2$ uniformly with a number of bits between 2 and 8 bits. While the quality of the correlation estimation degrades when the number of bits reduces, it is largely compensated by the reduction in bit-rate in the rate-distortion performance, as confirmed by Fig. 12. This means that the proposed correlation estimation is relatively robust to quantization so that it is possible to attain good rate-distortion performance by drastic quantization of the measurements.

Finally, we study the improvement offered by the robust data cost $\tilde{E}_d$ from Eq. (11) in Alg. 1 and Alg. 2 when the measurements have been compressed with a 2-bit uniform quantizer and an Arithmetic coder. We use the optimization toolbox based on CVX [2] in order to solve the optimization problem described in Eq. (10). Fig. 13 compares the modified Algorithms 1 and 2 with DSC based on disparity learning [4], a joint reconstruction with a block-based dictionary [3] and independent coding with JPEG-2000. The relative performance of the different schemes are maintained with the robust data cost $\tilde{E}_d$. However, the robust data cost permits to improve the quality of the image $\hat{I}_2$ around 1 dB especially at lower measurement rates. This gain can be observed by comparing the rate-distortion performance to the results given in Fig. 7 where the algorithms do not use the robust data cost term. For example, by comparing Fig. 7(b) and Fig. 13(b) we could see that for Plastic image set at bit-rate 0.6 bpp, the Alg. 2 improves quality of image $\hat{I}_2$ (approximately) from 29 dB to 30 dB when robust data term is used.

[2]available in http://cvxr.com/cvx/

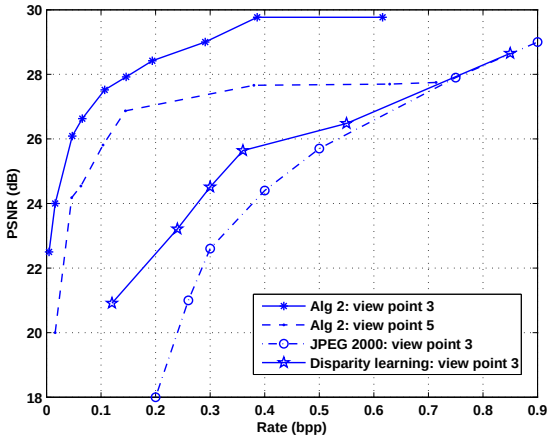


Fig. 10. Rate-distortion performance as a function of the correlation between images. Alg. 2 is used to reconstruct images at different viewpoints in the Sawtooth dataset, while the image at viewpoint 1 is selected as a reference image.

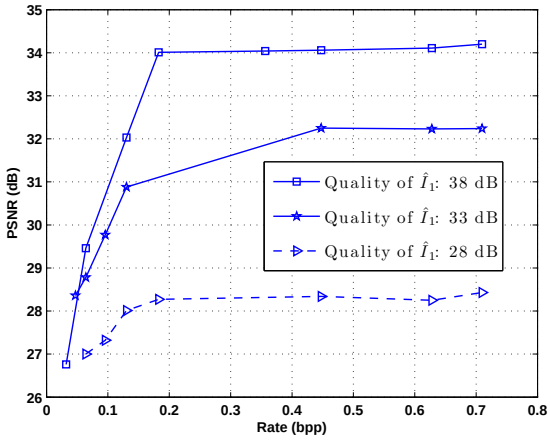


Fig. 11. Rate-distortion performance with Alg. 2 for reconstructing $\hat{I}_2$ as a function of the quality of the reference image $\hat{I}_1$ (resp. 28 dB, 33 dB and 38 dB) in the Plastic image set.

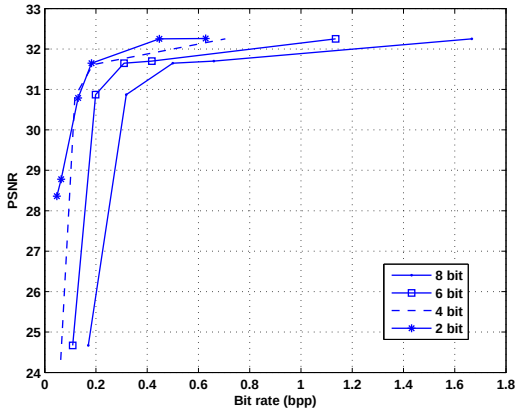
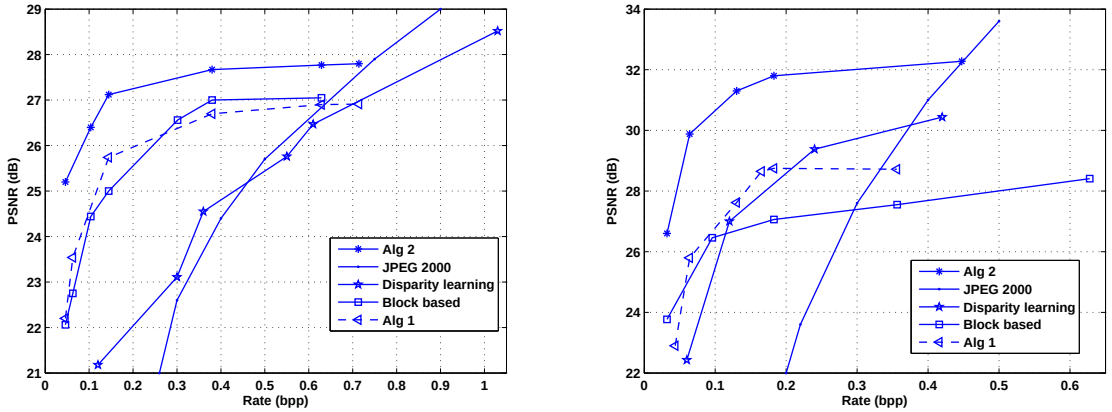


Fig. 12. Rate-distortion performance in reconstructed image $\hat{I}_2$ with Alg. 2 for different quantization rates in the Plastic image set.



(a) **(b)**

Fig. 13. Rate-distortion performance of Alg. 1 and Alg. 2 modified with a robust data cost term in the reconstruction of the image I_2. (a) Sawtooth image set (b) Plastic image set.

C. Distributed video coding

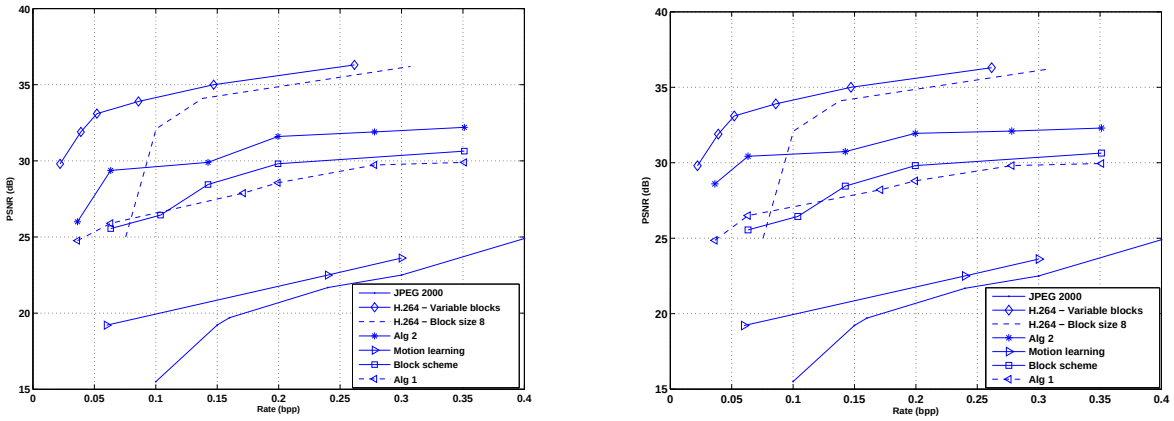
We study now the performance of the proposed algorithms in distributed video coding applications. The experiments are similar to the multi-view imaging framework above, except that the correlation estimation relates to motion estimation instead of disparity computation. We built the image set using the frames 2 and 3 of the Foreman sequence. The frame 2 is selected as the reference image I_1, and is approximated to a quality of approximately 45 dB in the joint decoder. We used the same dictionary described in the previous section for approximating the image $\hat{I}_1$. For this particular data set, we approximate $\hat{I}_1$ using $N = 60$ atoms. The measurements y_2 are compressed using a two-bit uniform quantizer and an Arithmetic coder. The search window size is $\delta t_x = \delta t_y = 4$ pixels for both the translational components t_x and t_y.

$(a) MSE: 80$ $(b) MSE: 101$ $(c) MSE: 41$ $(d) MSE: 73$

Fig. 14. Comparison of the warped image $\hat{I}_2$ w.r.t. I_2 and I_1 for the Foreman image set: (a) $1 - |\hat{I}_2 - I_2|$ with Alg. 1 (b) $1 - |\hat{I}_2 - I_1|$ with Alg. 1 (c) $1 - |\hat{I}_2 - I_2|$ with Alg. 2 (d) $1 - |\hat{I}_2 - I_1|$ with Alg. 2. [3801 quantized linear measurements].

Fig. 14 first illustrates the accuracy of the motion information computed in Alg. 1 and Alg. 2 with 3801 quantized measurements (i.e., 15% measurement rate). It compares the image $\hat{I}_2$ reconstructed by warping the reference image, to respectively the original images I_2 and I_1. We see that the warped image is closer to I_2 than I_1, which confirms the benefit of the motion estimation in the joint decoder. We further observe that the error denoted by black pixels is reduced significantly in the face region due to the good estimation of the motion field in smooth area. Similarly to the multi-view experiments, the motion around sharp edges is however not perfectly captured due to choice of the dictionary that does not include very thin geometrical patterns. Finally, we see that Alg. 2 provides a better estimation due to the benefit of the measurement consistency term E_t.

We further study the rate-distortion performance of the proposed algorithms in the reconstruction of the image $\hat{I}_2$. We compare the performance to different state-of-the-art solutions in distributed or joint video coding. First, we provide the performance of a DSC scheme (i.e., *Motion learning*) based on motion learning [5], using the similar experimental setup demonstrated in the previous section and a reference image $\hat{I}_1$ of 45 dB for a fair comparison. In addition, we implement Alg. 2 with a different dictionary that is built on blocks of the reference image, similarly to [3] (denoted as *Block-based* in the figures). We also compare to an independent encoding of the image I_2 with JPEG-2000. For the sake of completeness, we further provide results of a joint video encoding solution based on H.264 with an IP encoding structure (i.e., a GOP size of 2). We again encode the reference I frame (I_1) at a quality of 45 dB, and we vary the quantization parameter for the P frame (I_2) to build the rate-distortion characteristics. We consider two different settings in the H.264 motion estimation, which is performed with



(a) **(b)**

Fig. 15. Rate-distortion performance in reconstructing the second image $\hat{I}_2$ in Foreman data set with the Alg. 1 and Alg. 2 from measurements quantized on 2 bits. Comparisons with state-of-the-art coding solutions in joint and distributed video coding. The proposed algorithms are implement without (a) and with (b) robust data cost term.

variable and fixed macroblock size. Fig. 15 illustrates the rate-distortion performance of the different schemes and confirms the results that have been observed in the previous section. First, we see that the measurement consistency term E_t in Alg. 2 greatly improved the performance of our motion estimation algorithm. Similarly, the benefit of the robust data term in our algorithms is mostly visible at low rate, when we compare Fig. 15 (a) and Fig. 15 (b). Then, it is clear that our proposed solutions outperform independent coding since it exploits the correlation between images. It also outperforms DSC solution based on motion learning due a better model of the geometric correlation. The correlation estimation with block-based dictionary is less efficient than the estimation with a dictionary of geometric atoms. Finally, we see that the joint encoding with H.264 is clearly better than all distributed coding solutions. However, our algorithm is able to compete at low bit rate with H.264 based on a fixed block-size motion estimation, which is certainly an interesting and promising result.

VII. Conclusions

In this paper we have presented a framework for the distributed representation of image pairs with quantized linear measurements, along with joint reconstruction algorithms that exploit the geometrical correlation between images. We have a proposed a correlation model based on local transformations of geometric patterns that are present in the sparse representation of images. The motion or disparity information in distributed video coding and respectively multi-view imaging can then be estimated from the pairs of geometric features in different images. We propose a regularized optimization problem in order to identify the geometrical transformations that result in smooth motion or disparity fields between a reference and a predicted image. We have proposed a low complexity algorithms to the correlation estimation problem, which offers an effective trade-off between complexity and accuracy of the solution. In addition, we have proposed an improved reconstruction solution by image warping, where the image transformations are estimated in order to be consistent with the compressed measurements in the predicted image. Experimental results demonstrate that the proposed methodology provides a good estimation of dense disparity or motion fields in different natural image datasets. We also show that our geometry-based correlation model is more efficient than block-based correlation models. Finally, the consistent reconstruction constraints prove to offer improved reconstruction quality, such that the proposed algorithm outperform JPEG-2000 and DSC schemes in terms of rate-distortion performance. It certainly provides an interesting alternative to distributed image processing applications due to its effective framework based on geometry, which is the main characteristic of natural images.

References

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, pp. 1289–1306, 2006.
[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, pp. 489–509, 2006.
[3] J. P. Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. Picture Coding Symposium*, 2009.
[4] D. Varodayan, Y. C. Lin, A. Mavlankar, M. Flierl, and B. Girod, "Wyner-ziv coding of stereo images with unsupervised learning of disparity," in *Proc. Picture Coding Symposium*, 2007.
[5] D. Varodayan, D. Chen, M. Flierl, and B. Girod, "Wyner-ziv coding of video with unsupervised motion vector learning," *EURASIP Signal Processing: Image Communication*, vol. 23, pp. 369–378, 2008.
[6] E. J. Candes and J. Romberg, "Practical signal recovery from random projections," in *Proc. SPIE Computational Imaging*, 2005.

[7] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[8] S. Mun and J. Fowler, "Block compressed sensing of images using directional transforms," *Proc. IEEE International Conference on Image Processing*, 2009.

[9] L. Gan, T. T. Do, and T. D. Tran, "Fast compressive imaging using scrambled hadamard ensemble," in *Proc. European Signal and Image Processing Conference*, 2008.

[10] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *Proc. European Signal and Image Processing Conference*, 2008.

[11] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Proc. Picture Coding Symposium*, 2009.

[12] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. IEEE International Conference on Image Processing*, 2008.

[13] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Asilomar Conference on Signal System and Computing*, 2005.

[14] ——, "Universal distributed sensing via random projections," in *Proc. Information Processing in Sensor Networks*, 2006.

[15] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009.

[16] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proc. IEEE International Conference on Image Processing*, 2009.

[17] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 586–597, 2007.

[18] M. Trocan, T. Maugey, J. E. Fowler, and B. Pesquet-Popescu, "Disparity-compensated compressed-sensing reconstruction for multiview images," *Proc. IEEE International Conference on Multimedia and Expo*, 2010.

[19] M. Trocan, T. Maugey, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Multistage compressed-sensing reconstruction of multiview images," *Proc. IEEE International workshop on Multimedia Signal Processing*, 2010.

[20] I. Tosic and P. Frossard, "Geometry based distributed scene representation with omnidirectional vision sensors," *IEEE Trans. Image Processing*, vol. 17, pp. 1033–1046, 2008.

[21] O. D. Escoda, G. Monaci, R. M. Figueras, P. Vandergheynst, and M. Bierlaire, "Geometric video approximation using weighted matching pursuit," *IEEE Trans. Image Processing*, vol. 18, pp. 1703–1716, 2009.

[22] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Information Theory*, vol. 54, pp. 2210–2219, 2006.

[23] P. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Proc. International Conference on Information Science and Systems*, 2008.

[24] A. Schulz, L. Velho, and E. A. B. da Silva, "On the empirical rate-distortion performance of compressive sensing," in *Proc. IEEE International Conference on Image Processing*, 2009.

[25] L. Jacques, D. Hammond, and M. Fadili, "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine," *accepted to IEEE Trans. on Information Theory*.

[26] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 149–152, 2010.

[27] A. Fletcher, S. Rangan, and V. Goyal, "On the rate-distortion performance of compressed sensing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007.

[28] W. Dai, H. V. Pham, and O. Milenkovic. (2009) Distortion-rate functions for quantized compressive sensing. [Online]. Available: http://arxiv.org/abs/0901.0749

[29] J. Sun and V. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE International Symposium on Information Theory*, 2009.

[30] G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[31] R. M. Figueras, P. Vandergheynst, and P. Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Trans. Image Processing*, vol. 15, pp. 726–739, 2006.

[32] O. Veksler, "Efficient graph based energy minimization methods in computer vision," Ph.D. dissertation, Cornell University, 1999.

[33] P. Jost, P. Vandergheynst, and P. Frossard, "Tree-based pursuit: Algorithm and properties," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4685–4697, 2006.

[34] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automatic Control*, vol. 37, pp. 332–341, 1992.

[35] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense stereo," *International Journal on Computer Vision*, vol. 47, pp. 7–42, 2002.

[36] J. Maryak and D. Chin, "Global random optimization by simultaneous perturbation stochastic approximation," *IEEE Trans. on Automatic Control*, vol. 53, pp. 780–783, 2008.

[37] J. Whitney, S. Hill, D. Wairia, and F. Bahari, "Comparison of the spsa and simulated annealing algorithms for the constrained optimization of discrete non-separable functions," in *Proc. of American contol conference*, 2003.

[38] J. C. Spall, "Stochastic optimization: Stochastic approximation and simulated annealing," *Encyclopedia of Electrical and Electronics Engineering*, vol. 20, pp. 529–542, 1999.

[39] J. C. Spall, S. D. Hill, and D. R. Stark, "Theoretical framework for comparing several popular stochastic optimization approaches," in *Proc. of Americal Control conference*, 2002.

[40] L. Gerencser, S. Hill, Z. Vago, and Z. Vincze, "Discrete optimization, spsa, and markov chain monte carlo methods," in *Proc. of the American Control Conference*, 2004.

[41] O. Brooks, "Solving discrete resource allocation problems using the simultaneous perturbation stochastic approximation (spsa) algorithm," in *Proc. of the Spring Simulation Multiconference*, 2007.